Extreme disorder in an ultra-high-affinity protein complex

Alessandro Borgia[1†]*, Madeleine B. Borgia[1†], Katrine Bugge[2†], Vera M. Kissling[1], Pétur O. Heidarsson[1], Catarina B. Fernandes[2], Andrea Sottini[1], Andrea Soranno[1,3], Karin J. Buholzer[1], Daniel Nettels[1], Birthe B. Kragelund[2]*, Robert B. Best[4]*, Benjamin Schuler[1,5]*

[1]Department of Biochemistry, University of Zurich, Zurich, Switzerland

[2]Structural Biology and NMR Laboratory, The Linderstrøm-Lang Centre for Protein Science and Integrative Structural Biology at University of Copenhagen (ISBUC), Department of Biology, University of Copenhagen, Copenhagen, Denmark

[3]Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, Missouri, 63110, USA

[4]Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, USA

[5]Department of Physics, University of Zurich, Zurich, Switzerland

Supporting Information Methods

Protein preparation. Human prothymosin α (ProTα) was prepared and labeled as described previously[29]. For a complete list of all protein variants, labeling positions, and fluorophores used for single-molecule FRET, see Extended Data Table 1. The correct molecular mass of all protein variants and labeled constructs was confirmed by mass spectrometry.

For experiments using wild-type human linker histone H1.0 (H1), recombinant protein was used (New England Biolabs, cat.# M2501S). For the production of labeled H1 variants and wild-type H1 for NMR, the cDNA of the human HIF0 gene (UniProt P07305) was cloned into a modified version of the pRSET vector[52]. In this plasmid, the N-terminal hexahistidine tag and thrombin cleavage site were removed and replaced by a C-terminal hexahistidine tag and thrombin cleavage site (with sequence GGPRGSRGSHHHHHH) to enable purification of full-length H1 free of degradation products. Cysteine mutations were introduced via site-directed mutagenesis to allow for labeling with fluorescent dyes via maleimide coupling (see Extended Data Table 1 for a complete list of variants). All H1 variants were expressed in *E. coli* C41 cells and terrific broth medium at 37˚C, induced with 0.5 mM isopropylthiogalactopyranoside (IPTG) at an OD_{600} of ~0.6, and grown for 3 further hours. Cell pellets were harvested and resuspended in denaturing buffer (6 M guanidinium chloride (GdmCl) in phosphate-buffered saline (PBS, 10 mM sodium phosphate pH 7.4, 137mM NaCl, 2.7mM KCl), the soluble fraction was collected and applied to a Ni-IDA resin (ABT Beads, Spain) in batch. The resin was washed twice with 5 resin volumes of denaturing buffer including 25 mM imidazole, three times with 5 resin volumes of PBS including 25 mM imidazole, and the protein eluted with PBS including 250/500 mM imidazole. The protein was dialyzed against PBS, filtered, and its hexahistidine tag cleaved off with 5 U of thrombin (Serva) per milligram of H1, for 2 hours at room temperature. To remove uncleaved protein and the tag, the mixture was run through a HisTrap HP 5 ml column (GE Healthcare) in PBS including 25 mM imidazole. H1 was further purified using a Mono S ion exchange chromatography column (GE Healthcare), washed with 20 mM Tris (pH 8.0) including 200 mM NaCl, and eluted in 20 mM Tris (pH 8.0) buffer with a gradient from 200 mM to 1M NaCl. Finally, samples for labeling were reduced with 20 mM dithiothreitol and purified by reversed-phase chromatography on an HPLC with a Reprosil Gold C4 column with a gradient from 5% acetonitrile and 0.1% trifluoroacetic acid in aqueous solution to 100% acetonitrile. H1-containing fractions were lyophilized and resuspended in degassed 6 M GdmCl, 50 mM sodium phosphate buffer (pH 7.0). For double labeling, both dyes (dissolved in dimethylsulfoxide) were added to the protein in a 1:1:1 molar ratio;

for single labeling, dye was added at a 0.7:1 molar ratio of dye to protein. Reactions were incubated at room temperature for 2 hours, and stopped by adding 20 mM dithiothreitol. Products were purified by reversed-phase HPLC, and the correct mass of all labeled proteins confirmed by mass spectrometry (see Extended Data Fig. 8 for an example). Lyophilized labeled protein was dissolved in 8 M GdmCl and stored at -80 °C.

For NMR experiments, H1, ProTα, and the H1 globular domain (GD, Extended Data Table 1) were unlabeled or uniformly labeled with ^{15}N, or ^{15}N and ^{13}C by growing cells in LB media or in M9 minimal medium containing $^{15}NH_4Cl$ as the sole source of nitrogen and [$^{13}C_6$]-glucose as the sole source of carbon, respectively, as described previously[6] and purified essentially as explained above. GD was expressed as a GST-fusion protein with a TEV protease site, and purified on a Glutathione Sepharose 4 fast flow column (GE Healthcare). The column was washed with 10 column volumes (CV) of PBS and the tagged protein eluted with 5 CV of elution buffer (50 mM Tris-HCl pH 8.0, 10 mM reduced glutathione). All fractions containing GST-GD were pooled and cleaved with TEV protease (100 μL of 0.5 mg/mL stock solution) overnight, and subsequently applied to a HiTrap SP FF 5 mL (GE Healthcare) with 50 mM sodium phosphate pH 9.0 and eluted with 50 mM sodium phosphate pH 9.0, 1 M NaCl over 25 CV. The protein containing fractions were applied to a Superdex 75 10/300 (GE Healthcare) in TBS buffer (10 mM Tris, 157 mM KCl, 0.1mM EDTA, pH 7.4) and further concentrated using an Amicon Ultra–15 centrifugal filter device (Millipore) with a molecular weight cutoff of 3 kDa. Protein concentrations of H1 and GD were determined by UV absorbance, while the concentration of ProTα was determined by BCA assay (Thermo Scientific).

NMR spectroscopy. To minimize amide exchange, all NMR spectra were acquired at 283 K unless otherwise specified on a Varian INOVA 800 MHz (^1H) spectrometer with a room temperature probe or Bruker AVANCE III 600- or 750 MHz (^1H) spectrometers equipped with cryogenic probes. Free induction decays were transformed and visualized in NMRPipe[53] or Topspin (Bruker Biospin) and analyzed using the CcpNmr Analysis software[54]. Assignments of backbone nuclei of ^{13}C,^{15}N-labeled ProTα in the unbound state (0.1 mM ^{13}C,^{15}N-labeled ProTα, TBS buffer, 10 % D_2O (v/v), 0.7 mM 4,4-dimethyl-4-silapen-tane-1-sulfonic acid (DSS)) and at sub-saturating concentration (1:0.8 molar ratio) of H1 (0.1 mM ^{13}C,^{15}N-labeled ProTα, 0.08 mM H1, TBS buffer, 10 % D_2O (v/v), 0.7 mM DSS) were performed manually from the analysis of ^1H,^{15}N-HSQC-, HNCACB-, CBCA(CO)NH-, HN(CO)CA-, HNCO- and HN(CA)NNH spectra acquired with non-uniform sampling[55] using standard pulse sequences. At saturating concentrations of H1, backbone resonances of ProTα became too weak for successful assignments. Proton chemical shifts were referenced internally to DSS at 0.00 ppm, with heteronuclei referenced by relative gyromagnetic ratios. The content of transient structure in ProTα was evaluated for each state from secondary C^α-chemical shifts assigned in the free form and at 80% saturation of H1 using a random coil reference set for IDPs[25]. In both states, three transiently populated α-helices were identified: residues S9-E19 (~10% populated), A82-T86 (~13% populated), and V99-K102 (~18% populated). The populations of the transient α-helices were estimated from the average secondary chemical shift (SCS)-value of the residues of the transient helices divided by 2.8 ppm ($SCS_{C\alpha}$ value expected for a fully populated α-helix)[56] and were very similar in the free and bound states. ^1H,^{15}N-HSQC spectra of ^{15}N-labeled H1 (40 μM) were recorded in the absence and presence of ProTα (40 μM). ^1H,^{13}C-HSQC- and/or ^1H,^{15}N-HSQC spectra were acquired on four different sequential titrations: addition of up to 44 μM H1 to 11 μM ^{15}N-labeled ProTα, addition of up to 140 μM GD to 20 μM ^{15}N-labeled ProTα, addition of up to 400 μM ProTα to 100 μM ^{13}C,^{15}N-labeled H1-GS-6xH, and addition of up to 700 μM ProTα to 100 μM ^{13}C,^{15}N-labeled GD. Before each titration, the proteins were concentrated and dialyzed in the same beaker. Subsequently, the solution of labeled protein was split equally into two samples, to one of which the unlabeled titrant was added at the maximum concentration, and to the other the same volume of dialysis buffer. After acquisition of NMR spectra on the two samples, they were used to obtain titration points between the end points by sequentially mixing the sample of the complex into the free protein. All NMR titration data were recorded in TBS buffer, 10 % D_2O (v/v), 0.7 mM DSS. Binding-induced weighted chemical shift perturbations (CSPs) were calculated as the weighted Euclidean distance between the peaks using $|\gamma_N|/|\gamma_H| = 0.154$. Due to extensive resonance overlap of H^N, N, C^α and C^β resonances in the 2D- and 3D NMR spectra, assignments of backbone nuclei were not possible for the Glu

repeat region from E62-E67. Nonetheless, spin systems displaying resonances consistent with Glu residues with Glu neighbors could be identified, and by exclusion were assigned to be part of the E62-E67 Glu repeat. For three of these systems, amide backbone peaks could be confidently tracked in the titration of ^{15}N-labeled ProTα with H1. The intensity ratios and weighted CSPs of the three Glu amide backbone peaks upon addition of equimolar H1 were calculated and the average value used to represent the Glu repeat region in Figs. 3f,g. Peaks from the remaining Glu residues were present in the spectra of both free and bound states of ProTα, but could not be followed unambiguously during the titrations.

The hydrodynamic radii, R_H, of ProTα alone and at saturating concentrations of H1 or GD were determined from a series of ^{1}H,^{15}N-HSQC spectra with preceding pulse-field gradient stimulated-echo longitudinal encode decode (PG-SLED) diffusion filter[57] and with the gradient strength increasing linearly from 0.963-47.2 G/cm. To determine the diffusion coefficients, D, the decay curves of the amide peaks were plotted against the gradient strength and fitted in Dynamics Center (Bruker) using $I = I_0 \exp(-Dx^2\gamma^2\delta^2(\Delta - \delta/3) \cdot 10^4)$, with I being the intensity of the NMR signal at the respective gradient strength, I_0 the intensity without applied gradient, x the gradient strength in G/cm, γ = 26752 rad/(Gs), δ = 3 ms, Δ = 250 ms. R_H was calculated from the diffusion coefficient using the Stokes-Einstein relation, $R_H = k_B T/(6\pi\eta D)$, with η being the viscosity of water at 283 K.

Longitudinal (T_1) and transverse (T_2) ^{15}N-relaxation times were determined from two times two series of ^{1}H,^{15}N-HSQC spectra with varying relaxation delays using the pulse sequence of reference [58], employing pulsed-field gradients to suppress solvent resonances. The series were recorded on free ^{15}N-ProTα and on ^{15}N-ProTα with saturating concentrations of H1 at 800 MHz (^{1}H), using eight (10 ms, 100 ms, 300 ms, 500 ms, 700 ms, 1100 ms, 1300 ms, 1500 ms) and seven (50 ms, 90 ms, 130 ms, 190 ms, 230 ms, 390 ms, 490 ms) different relaxation delays for T_1 and T_2, respectively, plus triplicate measurements. The relaxation decays were fitted to single exponentials and relaxation times determined using the CcpNmr Analysis software[54].

Single-molecule fluorescence spectroscopy. Single-molecule measurements were performed using either a custom-built confocal instrument[59] or a MicroTime 200, both equipped with a HydraHarp 400 counting module (Picoquant, Berlin, Germany). The donor dye was excited with light from a 485-nm diode laser (LDH-D-C-485, PicoQuant) at an average power of 100 μW at the sample. The laser was operated in continuous-wave mode or in pulsed mode with alternating excitation of the dyes, achieved using pulsed interleaved excitation[60] (PIE). The wavelength range used for acceptor excitation was selected with a z582/15 band pass filter (Chroma) from the emission of a supercontinuum laser (EXW-12 SuperK Extreme, NKT Photonics, Denmark) driven at 20 MHz, which triggers (interleaved) pulses from the 485-nm diode laser used for donor excitation. Emitted photons were collected by the microscope objective (Olympus UplanApo 60×/1.20W), focused onto a 100 μm pinhole, and then separated into four channels with a polarizing beam splitter and two dichroic mirrors (585DCXR, Chroma). Emission was additionally filtered by bandpass filters (ET525/50M and HQ650/100, Chroma) before being focused onto one of four single-photon avalanche detectors (Optoelectronics SPCM AQR-15, PerkinElmer, Wellesley, MA, or τ-SPADs, PicoQuant, Germany).

FRET efficiency histograms of doubly labeled ProTα and H1 were acquired on samples with concentrations of labeled protein between 10 and 100 pM. For intermolecular measurements, up to 500 pM of acceptor-labeled protein were used to ensure saturation of binding. Measurements were performed in TBS buffer (165 mM ionic strength) or in an analogous buffer with higher ionic strength (adjusted by increasing the KCl concentration, as noted), in the presence of 140 mM β-mercaptoethanol (Sigma) for photoprotection[61] and 0.01% Tween 20 (Pierce) to minimize surface adhesion[62]. To avoid the pronounced interaction of H1 with glass surfaces, more inert polymer sample chambers (μ-Slide, ibidi, Germany) were used throughout. Transfer efficiencies were obtained from $E = n_A/(n_A + n_D)$, where n_D and n_A are the numbers of donor and acceptor photons, respectively, in each burst, corrected for background, channel crosstalk, acceptor direct excitation, differences in quantum yields of the dyes, and detection efficiencies[62]. Even in cases where PIE was insufficient to completely eliminate the donor-only contribution to the signal

(Fig. 3i), the population at zero transfer efficiency was sufficiently well separated from the FRET population that the reliability of the transfer efficiencies was not affected. Fluorescence anisotropy values were determined for all labeling positions via polarization-sensitive detection in the single-molecule instrument[28,63], and were between 0.04 and 0.16 for the monomeric proteins, and between 0.08 and 0.22 in the complex, indicating sufficiently rapid orientational averaging of the fluorophores to justify the approximation $\kappa^2 \approx 2/3$ used in Förster theory[64].

The low fluorescence anisotropy values, the consistency of the FRET and NMR results, and the self-consistency of a large number of labeling positions suggest that the fluorophores do not entail a severe perturbation of the interaction between ProTα and H1. However, to assess the effect of fluorophore labeling in more detail, we tested how different dye pairs and labeling positions influence the affinity between ProTα and H1 and the inferred inter-dye distances (Extended Data Table 2). In view of the high net charge of the proteins, alternative fluorophores with a net charge different from Alexa 488 and 594 (both net charge -2) were chosen: Cy3B (GE Healthcare Life Sciences; zwitterionic with zero net charge), Abberior STAR 635 (Abberior GmbH; zwitterionic with zero net charge), and Atto550 and Atto647N (ATTO-TEC; both net charge +1). The K_D values for the respective binding partner were between 1.0 nM and 3.5 nM (at 205 mM ionic strength to simplify quantitation) for all labeling positions and dye pairs, corresponding to an energetic perturbation of binding by at most ~1 k_BT. To test for the effect of the fluorophores on the inferred distances, we recorded single-molecule transfer efficiency histograms of ProTα 56-110 labeled with Cy3B/Abberior*635 and Atto550/Atto647N, and of H1 104-194 labeled with Cy3B/Abberior Star 635, both with and without the respective unlabeled binding partner present. The resulting transfer efficiency values yielded root mean square interdye distances consistent with those inferred from measurements with Alexa 488/594 (assuming a Gaussian chain distribution of inter-dye distances[29] and an experimental uncertainty of ±0.05 for the transfer efficiency due to instrument calibration for the different dye pairs).

Fluorescence lifetime analysis. The comparison of ratiometric transfer efficiencies with the mean fluorescence lifetimes of donor and acceptor provides a further diagnostic for the presence of a broad distance distribution rapidly sampled during the time of a fluorescence burst[28,33,34]. Average lifetimes were estimated by using the mean donor ($\langle t_D \rangle$) and acceptor ($\langle t_A \rangle$) arrival times of the respective photons in a burst relative to the exciting laser pulse, and were combined with transfer efficiencies in a two-dimensional plot (Extended Data Fig. 5), where for each burst, $\tau_D^D/\tau_D^0 = \langle t_D \rangle/\tau_D^0$ and $\tau_D^A/\tau_D^0 = (\langle t_A \rangle - \tau_A^0)/\tau_D^0)$ were calculated. Here, τ_D^0 is the intrinsic donor lifetime in the absence of the acceptor, and τ_A^0 is the intrinsic acceptor lifetime. For a single, fixed inter-dye distance (and thus transfer efficiency, E), one finds $\tau_D^D/\tau_D^0 = \tau\,{}_D^A/\tau_D^0 = 1\text{-}E$, as illustrated by the diagonal line in Extended Data Fig. 5.

Nanosecond fluorescence correlation spectroscopy (nsFCS). Data for nsFCS were acquired at a concentration of ~100 pM of the protein carrying the donor (or both donor and acceptor) and an excess of the partner (either unlabeled or acceptor-labeled) to saturate binding. Donor and acceptor fluorescence emission (upon continuous-wave excitation at 485 nm) from the subpopulation corresponding to the H1-ProTα complex in a transfer efficiency histogram was correlated with a binning time of 1 ns. To avoid effects of detector dead times and afterpulsing on the correlation functions, the signal was recorded with two detectors each for donor and acceptor and cross-correlated between detectors[34,35]. Autocorrelation curves of acceptor and donor channels and crosscorrelation curves between acceptor and donor channels were computed from the measurements and analyzed as described previously[34,65]. Briefly, auto- and crosscorrelation curves were fitted over a time window of 2.5 μs with

$$g_{ij}(\tau) = 1 + \frac{1}{N}\left(1 - c_{ab}e^{-|\tau|/\tau_{ab}}\right)\left(1 + c_{cd}e^{-|\tau|/\tau_{cd}}\right)\left(1 + c_T e^{-|\tau|/\tau_T}\right), \qquad i, j = A, D,$$

where i and j correspond to donor or acceptor fluorescence emission; N is the effective mean number of molecules in the confocal volume; c_{ab}, τ_{ab}, c_{cd} and τ_{cd} are the amplitudes and time constants of photon antibunching and chain dynamics, respectively; and c_T and τ_T refer to the triplet blinking component on the

microsecond timescale. Distance dynamics result in a characteristic pattern of the correlation functions based on donor and acceptor emission, with a positive amplitude in the autocorrelations ($c_{cd} > 0$) and a negative amplitude in the crosscorrelation ($c_{cd} < 0$), but with identical decay times. All three correlation curves were thus fitted globally with the same values of τ_{cd}. Independent values of c_{cd}, c_{ab}, τ_{ab}, τ_T, and c_T were used as free fit parameters for each correlation curve. τ_{cd} was converted to the reconfiguration time of the chain, τ_r, as described in Gopich et al.[65], by assuming that chain dynamics can be modelled as a diffusive process in the potential of mean force derived from the sampled inter-dye distance distribution $P(r)$.[35,65] In view of the good agreement between the transfer efficiencies observed experimentally and in the simulations, we employed $P(r)$ obtained from the simulations for the respective pairs of labeling sites (intra- or intermolecular). Note that this conversion does not entail a large change in timescale, and τ_{cd} and τ_r differ by less than 20% in all cases investigated here, depending on the average distance relative to the Förster radius[65]. The correlation functions shown in Fig. 3a-d were normalized to 1 at their respective values at 0.5 μs to facilitate direct comparison.

Two-focus fluorescence correlation spectroscopy (2f-FCS). 2f-FCS measurements of Alexa 594-labeled ProTα were performed at 295 K on a MicroTime 200 confocal microscope equipped with a differential interference contrast prism. Alexa 594 was excited alternatingly with two orthogonally polarized laser beams: one beam with a wavelength range centered at 582 nm, selected with a z582/15 band pass filter (Chroma) from the emission of a supercontinuum fiber laser (EXW-12 SuperK Extreme, NKT Photonics, Denmark) driven at 20 MHz, triggers (interleaved) pulses from a second supercontinuum laser with wavelength-selected output at 585 ± 3 nm (Solea, PicoQuant, Germany), with a combined repetition rate of 40 MHz and a power of 15 μW per laser at the sample. The distance between the two foci, δ, was calibrated as described previously based on sample standards quantified under identical conditions using dynamic light scattering[30,66], yielding a δ of 490 ± 15 nm at λ_{ex} = 585 nm, corresponding to a systematic error of 3% of the calculated value of the hydrodynamic radius R_H. The concentration of labeled protein used in these experiments was ~4 nM in TBS buffer in the presence of 140 mM β-mercaptoethanol and 0.01% Tween 20. Translational diffusion coefficients were obtained from fits of the correlation functions[67] and converted to R_H using the Stokes-Einstein equation.

Analysis of binding isotherms. At ionic strengths of 200 mM and above, binding titrations of ProTα and H1 were hyperbolic and could be described well with a Langmuir-type isotherm, valid when the ligand concentration is sufficiently large compared to the analyte concentration, e.g. with H1 as the ligand and ProTα as the analyte:

$$\frac{c_{\mathrm{ProT\alpha\text{-}H1}}}{c_{\mathrm{ProT\alpha}}^{\mathrm{tot}}} = \frac{c_{\mathrm{H1}}^{\mathrm{tot}}}{K_{\mathrm{D}} + c_{\mathrm{H1}}^{\mathrm{tot}}},$$

where c_{X} indicates the concentration of species X, and $c_{\mathrm{X}}^{\mathrm{tot}}$ the total concentration of X. Below an ionic strength of about 200 mM, however, the affinity of H1 for the surface of the sample chambers in which the measurements were performed was so high that they noticeably competed with H1 binding by ProTα (note that the polymeric sample chambers used here already exhibit much lower affinity for H1 than glass surfaces, which are negatively charged). This results in a decrease in the effective H1 bulk concentration available for binding to ProTα and leads to a shift of the apparent midpoint of the titration to higher H1 concentrations and a distortion of the curve to a non-hyperbolic shape. To account for this effect, we need to take into account two coupled equilibria, one for the adsorption of H1 to surface binding sites, S, and one for H1 binding to ProTα:

$$\mathrm{H1} + \mathrm{S} \rightleftarrows \mathrm{H1S} \quad \text{and} \quad \mathrm{H1} + \mathrm{ProT\alpha} \rightleftarrows \mathrm{H1ProT\alpha},$$

with the dissociation constants

$$K_{\mathrm{D}}^{\mathrm{H1S}} = \frac{c_{\mathrm{H1}} \cdot \Gamma_{\mathrm{S}}}{\Gamma_{\mathrm{SH1}}} \quad \text{and} \quad K_{\mathrm{D}}^{\mathrm{H1\text{-}ProT\alpha}} = \frac{c_{\mathrm{H1}} \cdot c_{\mathrm{ProT\alpha}}}{c_{\mathrm{ProT\alpha\text{-}H1}}}, \qquad \text{Eqs. 1 and 2}$$

where c_{H1}, $c_{\mathrm{ProT\alpha}}$, and $c_{\mathrm{ProT\alpha\text{-}H1}}$ are the bulk concentrations of free H1, free ProTα, and complex, respectively. Γ_{S} and Γ_{SH1} are the surface concentrations (i.e., binding sites per area) of free binding sites and of binding sites occupied by H1, respectively. Correspondingly, the resulting total concentrations are

$$c_{\mathrm{H1}}^{tot} = c_{\mathrm{H1}} + c_{\mathrm{ProT\alpha\text{-}H1}} + \alpha\, \Gamma_{\mathrm{SH1}} \qquad \text{Eq. 3}$$

$$c_{\mathrm{ProT\alpha}}^{\mathrm{tot}} = c_{\mathrm{ProT\alpha}} + c_{\mathrm{ProT\alpha\text{-}H1}} \qquad \text{Eq. 4}$$

$$\Gamma_{\mathrm{S\,tot}} = \Gamma_{\mathrm{S}} + \Gamma_{\mathrm{SH1}} \qquad \text{Eq. 5}$$

Here, α is the surface-to-volume ratio of the sample well. Eqs. 1 to 5 were solved for the fraction of H1-bound ProTα using Mathematica (Wolfram Research) and the solution used to fit the titrations with full-length H1 and the H1 C-terminal fragment at 165 mM ionic strength (Fig. 2b) and full-length H1 at 185 mM ionic strength (Fig. 2c), with the adjustable parameters $K_D^{H1\text{-}ProT\alpha}$, K_D^{H1S}, and with the product $\alpha \cdot \Gamma_{SH1}$; $c_{\mathrm{ProT\alpha}}^{\mathrm{tot}}$ was fixed to the known value. The vertical error bars in Fig. 2b were estimated from five independent measurements. The horizontal error bars represent the pipetting errors estimated for the applied sequences of dilution steps. By additionally taking into account the uncertainty of the ProTα concentration, we obtain upper and lower bounds for the binding isotherms, which are displayed as shaded bands in Fig. 2b. Note that the K_{D} resulting for the full-length proteins at 165 mM ionic strength follows the trend expected from the measurements at higher ionic strength (Fig. 2c), validating the analysis. The weak association of additional monomers at high micromolar excess of binding partner was ignored in this analysis since it occurs in a different concentration regime. The dependence of the K_{D} on ion activity, a, (Fig. 2c) was analyzed using the formalism developed by Record and Lohman[51] according to the approximation $d \ln K_D / d \ln a \approx -\Delta n = 19 \pm 1$ (standard error of the fit), where -Δn corresponds to the number of anionic and cationic counter ions released upon association of the two proteins, and the activity was approximated by the ionic strength.

Circular dichroism spectroscopy. Far-UV CD measurements were carried out on a Jasco J-810 spectropolarimeter, using a 1 mm path length quartz cuvette. Wild type H1 and ProTα E56C samples were measured at a concentration of 5 µM in TBS and 5 mM β-mercaptoethanol at 20 °C. A total of 20 to 60 spectra per sample were recorded between 250 and 195 nm with 1-nm step size, averaged, and a buffer spectrum subtracted. The far-UV CD spectrum of the GD was recorded at 283 K from 260 to 198 nm with a scan speed of 20 nm/min, 10 accumulations and a response time of 2 s at a protein concentration of 10 µM in TBS and the buffer spectrum subtracted. To assess the thermal stability of the GD, thermal unfolding was monitored at 222 nm from 283 to 378 K at 1 K increments per minute. The ellipticity as a function of temperature was fitted with

$$\theta(T) = f_U(T)\theta_U(T) + (1 - f_U(T))\theta_N(T),$$

$$\text{where } f_U(T) = \left(1 + \exp\left(-\frac{\Delta H_m}{R}\left(\frac{1}{T} + \frac{1}{T_m}\right)\right)\right)^{-1}$$

is the fraction of unfolded GD; ΔH_m the enthalpy change of folding at the transition midpoint, and R the gas constant. $\theta_N(T)$ and $\theta_U(T)$ are linear baselines from the folded and unfolded states, respectively, as a function of absolute temperature T.

Binding kinetics of H1 and ProTα. Mixing experiments were carried out with an Applied Photophysics Pi*-180 stopped-flow spectrometer. A solution of ProTα doubly labeled with Alexa 488/594 (positions 56 and 110) at a concentration of 2.2 nM was mixed with a solution of unlabeled H1 at variable concentrations using a 1:10 mixing ratio. The increase in acceptor fluorescence emission resulting from the compaction of ProTα upon H1 binding (cf. Fig. 2a) was used to monitor the binding reaction by exciting at 436 nm with a 10-nm bandwidth using a HgXe lamp and recording fluorescence emission using a 580 nm long-pass filter. The buffer used was TBS in the presence of 0.01% Tween 20 to minimize surface adhesion of the proteins. For each final H1 concentration between 5 nM and 100 nM, at least 80 measurements were recorded and averaged.

Simulation methods. A coarse-grained model was used for both proteins, in which each residue is represented by a single bead centered on the α-carbon atom. The potential energy had the functional form:

$$V = \frac{1}{2}\sum_{i<N} k_b\left(d_i - d_i^0\right)^2 + \frac{1}{2}\sum_{i<N-1} k_\theta\left(\theta_i - \theta_i^0\right)^2 + \sum_{i<N-2}\sum_{n=1}^{4} k_{i,n}(1+\cos(n\phi_i - \delta_{i,n})) + \sum_{i<j}\frac{q_i q_j}{4\pi\epsilon_d\epsilon_0}\exp\left[-\frac{d_{ij}}{\lambda_D}\right] + \sum_{(i,j)\notin \text{nat}} 4\varepsilon_{pp}\left(\left(\frac{\sigma_{ij}}{d_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{d_{ij}}\right)^{6}\right) + \sum_{(i,j)\in \text{nat}} 4\varepsilon_{ij}\left(13\left(\frac{\sigma_{ij}}{d_{ij}}\right)^{12} - 18\left(\frac{\sigma_{ij}}{d_{ij}}\right)^{10} + 4\left(\frac{\sigma_{ij}}{d_{ij}}\right)^{6}\right)$$

The first two terms describe harmonic bond and angle energies, respectively, with force constants $k_b = 3.16 \times 10^5$ kJ.mol^{-1}nm^{-2} and $k_\theta = 6.33 \times 10^2$ kJ mol^{-1} rad^{-2}, and reference values d_i^0 and θ_i^0 taken from an extended backbone structure. The third term is a sequence-based statistical torsion potential taken from the Go model of Karanicolas and Brooks[38], which is applied to all residues, and the fourth term is a screened coulomb potential, with Debye screening length λ_D applied to all residues with non-zero charges q_i; ϵ_0 is the permittivity of free space and ϵ_d the dielectric constant, set here to 80. The fifth term is a generic short-range attractive potential applied to all residue pairs not identified as being part of the natively folded globular domain of H1. This interaction is characterized by a contact distance $\sigma_{ij} = (\sigma_i + \sigma_j)/2$, where $\sigma_{i,j}$ are the residue diameters (all ~6 Å) determined from residue volumes[37], and by a contact energy ε_{pp}, which is the same for all such non-native residue pairs. The final term is an attractive potential applied only to the residues identified as native in the folded histone domain. Which residues are considered native, and the values of the parameters σ_{ij} and ε_{ij} for native pairs are given by the Karanicolas and Brooks Go model[38]. For the electrostatic term, the charges are +1 for lysine and arginine, -1 for glutamate and aspartate, and +0.5 for histidine (to account for its pKa near 6). The screening length, or Debye length, λ_D is given by

$$\lambda_D = \left(\frac{\epsilon_d\epsilon_0 k_B T}{2N_A e^2 I}\right)^{1/2},$$

where k_B is the Boltzmann constant, T the temperature, N_A the Avogadro constant, e the elementary charge, and I the ionic strength in molar units. The variation of ionic strength only enters the model via the screening length. Although this treatment of electrostatics is very simplified, it is consistent with the level of coarse-graining in the rest of the model.

Thus there was only one free parameter to be determined, namely ε_{pp}; the same value was used for all interactions, whether inter- or intramolecular. We varied ε_{pp} in order to obtain an optimal agreement with all the FRET data. This optimal value was found to be $0.16\ k_BT$, or ~ 0.40 kJ/mol^{-1}. Langevin dynamics simulations were run at a temperature of 300 K, with a friction coefficient of 0.1 ps^{-1} and a time step of 10 fs for 20 μs for each run; the mass of each bead was that of the corresponding residue. Simulations of the bound complex were started either with the molecules separated, or in an initially contacting configuration. Results from either simulation were the same, neglecting the equilibration part of the simulation. We also tested the effect of variations of the model. Using a residue-independent value of 6 Å for σ_{ij} for all residue pairs did not appreciably change the results. Similarly, using a residue-specific short-range potential similar to that devised in the protein interaction model by Kim and Hummer[68] did not improve the agreement with experiment. However, a model with randomized or uniform charges (equal to the average) for the two proteins was unable to capture the important qualitative features of the data, in particular the difference in FRET efficiencies between the N- and C-termini of ProTα, and H1. This result emphasizes the dominant role of electrostatics in determining the properties of the complex.

We also considered whether the results may be influenced by the presence of the FRET chromophores and the linkers used to covalently attach them to the protein. We therefore ran an additional set of simulations, one for each labeling combination, in which we included an explicit, coarse-grained representation of the linkers. The linkers and dyes were approximated by 5 beads (for each dye+linker), in an unbranched chain, and with similar properties to the protein (bond lengths 3.8 Å, all bond angles 110°; the dihedral angle term was omitted). One end of the chain was bonded to the bead for the labeled residue. The motivation for the choice of 5 beads and protein-like geometry was the earlier finding that the effect of linkers on unfolded proteins can be accounted for by adding an extra 9-10 residues to the true number of residues separating the labeling positions[30,66]. The short-range interaction of the dyes with themselves and the protein was given by a Lennard-Jones term like that used for the other non-native interactions in the model, but the parameters were set so as to give only a short-range repulsion, with $\varepsilon = 0.001$ kJ/mol, and $\sigma = 6$ Å. Each chromophore carries a net charge of -2, which was included by adding a charge of -1 to each of the two beads furthest from the attachment point to the protein. Explicit simulations were run for each labeling combination considered in the paper.

Dissociation constants were estimated by umbrella sampling using the center of mass distance between the proteins as coordinate, with harmonic umbrellas spaced between 0 and 25 nm and a force constant of 10 kJ mol^{-1} nm^{-2}. The potential of mean force $F_{\mathrm{WHAM}}(r)$ between the proteins was reconstructed using weighted histogram analysis (WHAM)[69], and the effective pair potential $F_{\mathrm{eff}}(r)$ (Extended Data Fig. 6b) was obtained from $F_{\mathrm{eff}}(r) = F_{\mathrm{WHAM}}(r) + 2k_BT \log r$, where k_B is the Boltzmann constant and T the temperature. F_{eff} was shifted by a constant energy so that the interaction energy at large separations was zero. The dissociation constant K_{D} was calculated from

$$K_D^{-1} = 4\pi N_A \int_0^{r_b} \exp[-\beta F_{\mathrm{eff}}(r)] r^2 dr,$$

where r_b is the radius defining the maximum extent of the bound state (where $F_{\mathrm{eff}}(r)$ becomes non-zero), and $\beta = 1/k_BT$.

Conformations were initially analyzed using the clustering algorithm devised by Rodriguez and Laio[70], applied to the Hamming distances between the binary contact maps for different conformations (using a distance cut-off of 8 Å to define a contact). This algorithm identifies cluster centers as structures i with a high density of neighbors, ρ_i, (many structures at a short distance), but which have a large distance to the nearest structure with higher neighbor density, δ_i. The "decision graph" consists of plotting δ_i vs. ρ_i for all structures. Cluster centers should appear as points at the top right of the graph. The decision graph in this case (Extended Data Fig. 6a) shows only a single cluster. Other clustering algorithms also gave little evidence for distinct clusters, suggesting that all structures fall into a single very broad state. We therefore used principal component analysis (PCA) as a way of projecting out the structural variations. We used a set of coarse-grained inter-residue distances as the space in which to perform PCA, in which only every 5$^{\mathrm{th}}$ residue in the sequence was considered, and all pair distances between such residues were computed. We

obtained the principal components by diagonalization of the variance-covariance matrix of this set of distances. The first three components are represented as matrices in Extended Data Fig. 6c.

Data availability. The authors declare that all data supporting the findings of this study are available within the paper and its supplementary information files. The raw data are available from the corresponding authors upon reasonable request. The chemical shift assignments of ProTα alone and in complex with H1 have been deposited to the Biological Magnetic Resonance Bank under accession numbers 27215 and 27216, respectively.

Code availability. A custom module for Mathematica (Wolfram Research) used for the analysis of single-molecule fluorescence data is available upon request.

References

52 Scott, K. A., Steward, A., Fowler, S. B. & Clarke, J. Titin; a multidomain protein that behaves as the sum of its parts. *J. Mol. Biol.* **315**, 819-829 (2002).

53 Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**, 277-293 (1995).

54 Vranken, W. F. *et al.* The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59**, 687-696 (2005).

55 Orekhov, V. Y. & Jaravine, V. A. Analysis of non-uniformly sampled spectra with multi-dimensional decomposition. *Prog Nucl Magn Reson Spectrosc* **59**, 271-292 (2011).

56 Fedyukina, D. V. *et al.* Contribution of long-range interactions to the secondary structure of an unfolded globin. *Biophys. J.* **99**, L37-39 (2010).

57 Gibbs, S. J. & Johnson, C. S. A PFG NMR Experiment for Accurate Diffusion and Flow Studies in the Presence of Eddy Currents. *J. Magn. Reson.* **93**, 395-402 (1991).

58 Farrow, N. A. *et al.* Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. *Biochemistry* **33**, 5984-6003 (1994).

59 Benke, S. *et al.* The assembly dynamics of the cytolytic pore toxin ClyA. *Nat. Commun.* **6**, 6198 (2015).

60 Müller, B. K., Zaychikov, E., Bräuchle, C. & Lamb, D. C. Pulsed interleaved excitation. *Biophys. J.* **89**, 3508-3522 (2005).

61 Rasnik, I., McKinney, S. A. & Ha, T. Nonblinking and long-lasting single-molecule fluorescence imaging. *Nat. Methods* **3**, 891-893 (2006).

62 Schuler, B. Application of single molecule Förster resonance energy transfer to protein folding. *Methods. Mol. Biol.* **350**, 115-138 (2007).

63 Kellner, R. *et al.* Single-molecule spectroscopy reveals chaperone-mediated expansion of substrate protein. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 13355-13360 (2014).

64 Förster, T. Zwischenmolekulare Energiewanderung und Fluoreszenz. *Annalen der Physik* **6**, 55-75 (1948).

65 Gopich, I. V., Nettels, D., Schuler, B. & Szabo, A. Protein dynamics from single-molecule fluorescence intensity correlation functions. *J. Chem. Phys.* **131**, 095102 (2009).

66 Borgia, A. *et al.* Consistent View of Polypeptide Chain Expansion in Chemical Denaturants from Multiple Experimental Methods. *Journal of the American Chemical Society* **138**, 11714-11726 (2016).

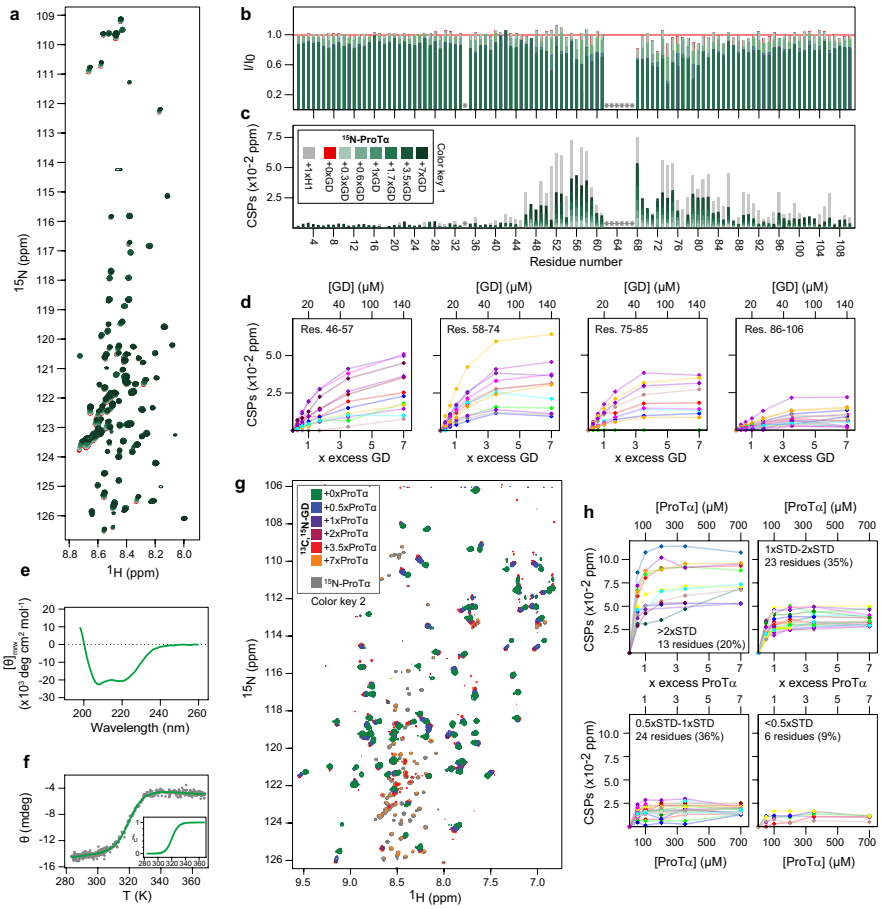
67 Dertinger, T. *et al.* Two-focus fluorescence correlation spectroscopy: A new tool for accurate and absolute diffusion measurements. *Chemphyschem* **8**, 433-443 (2007).

68 Kim, Y. C. & Hummer, G. Coarse-grained models for simulations of multiprotein complexes: application to ubiquitin binding. *J. Mol. Biol.* **375**, 1416-1433 (2008).

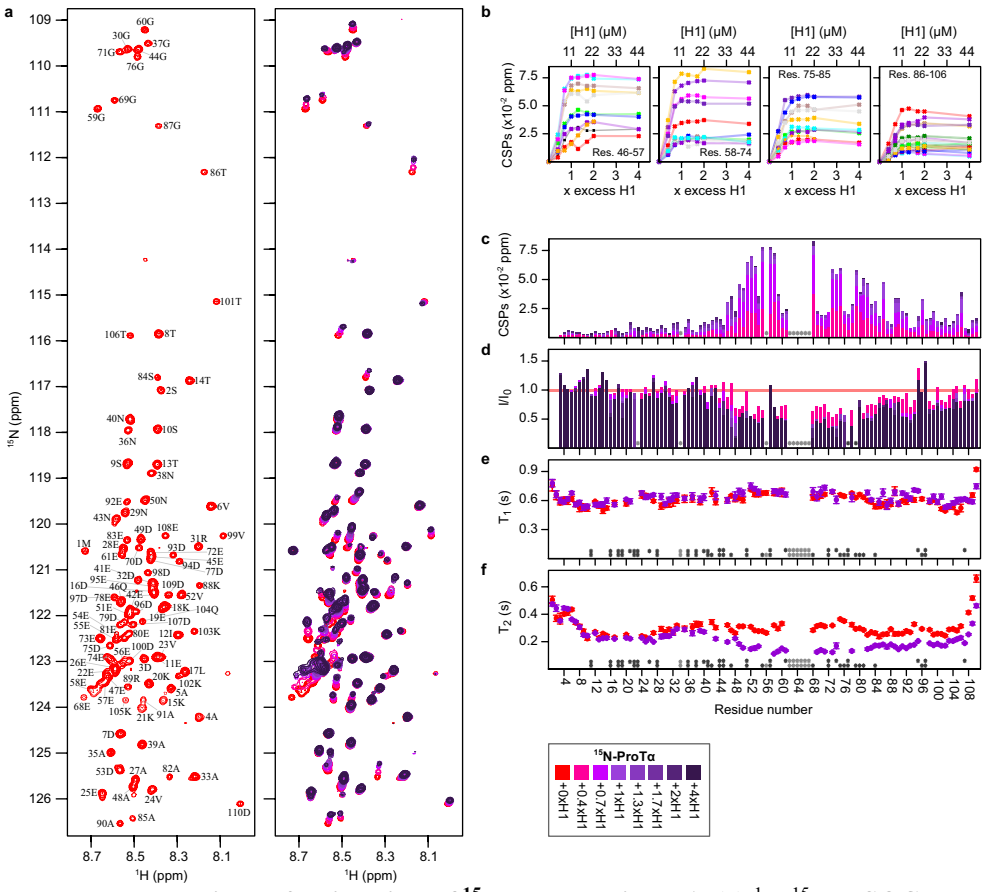
69 Kumar, S., Bouzida, D., Swendsen, R. H., Kollman, P. A. & Rosenberg, J. M. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comp. Chem.* **13**, 1011-1021 (1992).

70 Rodriguez, A. & Laio, A. Machine learning. Clustering by fast search and find of density peaks. *Science* **344**, 1492-1496 (2014).

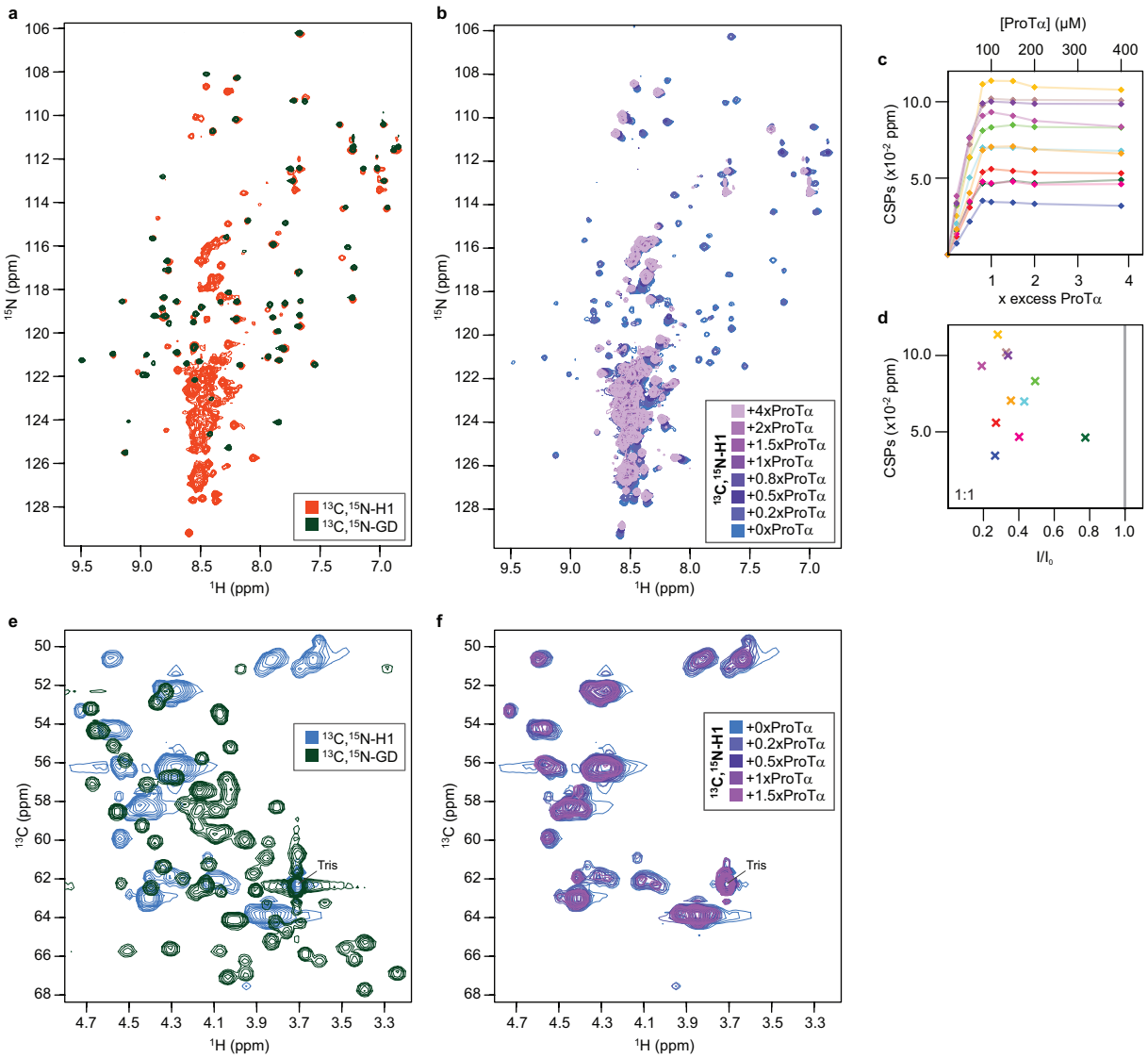
Extended Data Figures and Tables



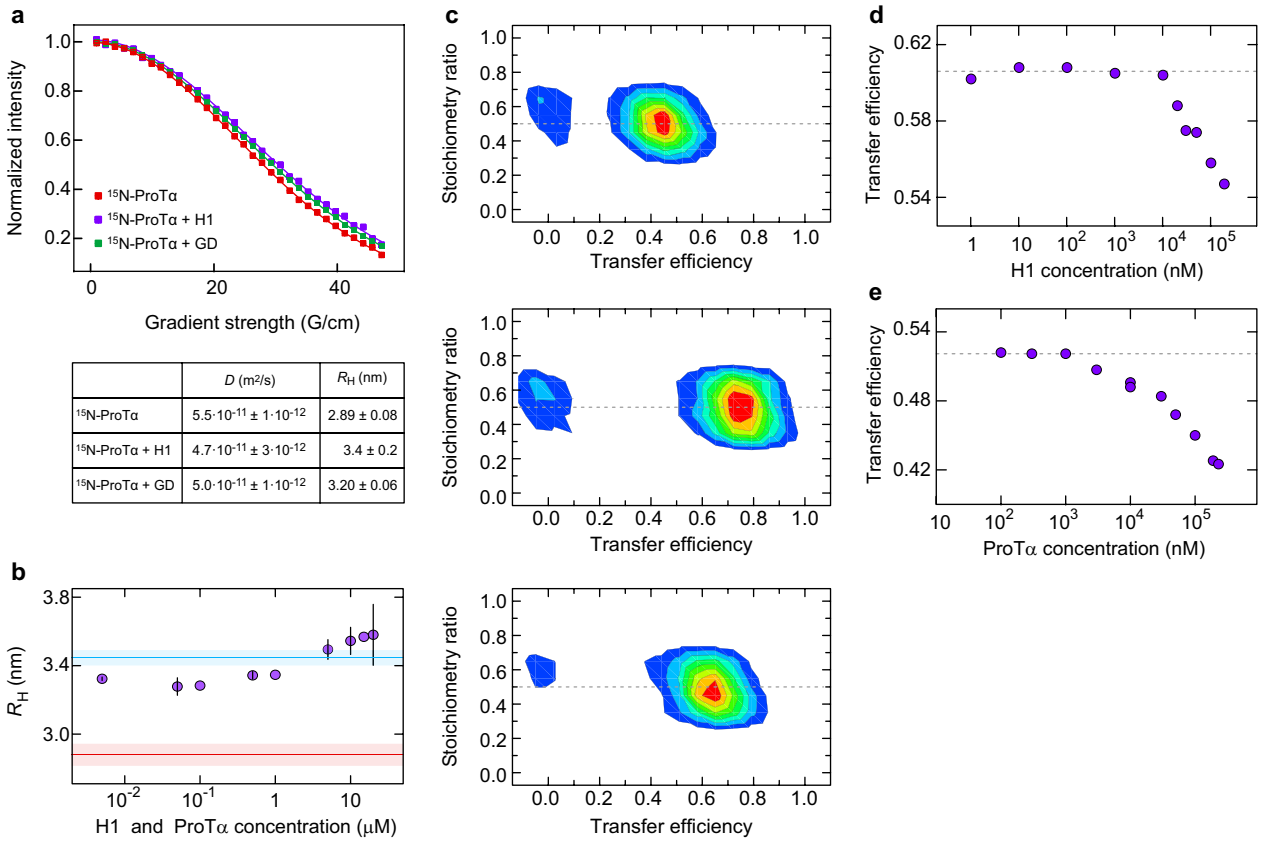
Extended Data Figure 1. Titrations of ProTα and Globular Domain (GD). (a) Titration of ^{15}N-ProTα with 0- to 7-fold molar addition of GD followed by ^{1}H,^{15}N-HSQC spectra. **(b)** Peak intensity ratios for assigned residues of ProTα relative to the free state induced by 0- to 1.7-fold molar addition of GD. **(c)** CSPs per residue of ProTα induced by 0- to 7-fold molar addition of GD. For comparison, CSPs of ProTα upon 1-fold molar addition of H1 are shown in grey. Panels **a-c** follow color key 1; light grey stars indicate prolines and unassigned residues. **(d)** ProTα CSPs plotted against concentration and times excess of GD relative to the free state for residues 46-106 upon 0- to 7-fold molar addition of GD. Colors used for discriminability. **(e)** Far-UV CD spectrum of GD. **(f)** Thermal denaturation of GD followed by the change in ellipticity at 222 nm (T_m = 320.5 ± 0.3 K, ΔH_m = -44 ± 2 kcal mol^{-1}). Insert: Fraction unfolded GD (f_u) as a function of temperature. **(g)** Titration of 100 μM ^{13}C,^{15}N-GD with 0- to 7-fold molar addition of ProTα followed by ^{1}H,^{15}N-HSQC spectra (color key 2). Peak intensities gradually decrease during the titration. At 3.5×- and 7× excess ProTα, natural abundance peaks of free ProTα appear (^{1}H,^{15}N-HSQC spectrum of ^{15}N-ProTα shown in grey for comparison). **(h)** CSPs of GD plotted against concentration and times excess of ProTα relative to the free state upon 0- to 7-fold molar addition of ProTα. A total of 66 (unassigned) amide backbone peaks were followed and grouped according to the standard deviation (STD) of the CSPs (1 STD = 0.0254 ppm). Of these, 55% had CSPs larger than 1 STD. Colors used for discriminability.



Extended Data Figure 2. Titration of ^{15}N-ProTα with H1. (a) ^{1}H,^{15}N-HSQC spectrum of 11 μM free ^{15}N-ProTα with residue labels (left) and titrated with 0- to 4-fold molar addition of H1 (right) (see color key). **(b)** Weighted backbone amide chemical shift perturbations (CSPs) of ProTα (residues 46-106) relative to the free state upon 0- to 4-fold molar addition of H1, plotted against concentration and times excess of H1. Colors used for discriminability. **(c)** CSPs and **(d)** peak intensity ratios for assigned residues of ProTα induced by 0- to 4-fold molar addition of H1 (for bar colors, see key). **(e)** Longitudinal ^{15}N relaxation times (T_1) of free (red) and H1-bound (purple) ^{15}N-ProTα. $\langle T_1 \rangle$ is 610 ms (free) and 636 ms (complex). **(f)** Transverse ^{15}N relaxation times (T_2) of free (red) and H1-bound (purple) ^{15}N-ProTα. $\langle T_2 \rangle$ is 302 ms (free) and 217 ms (complex). In c-f, light grey stars indicate prolines and unassigned residues, dark grey stars overlap and/or insufficient data quality.

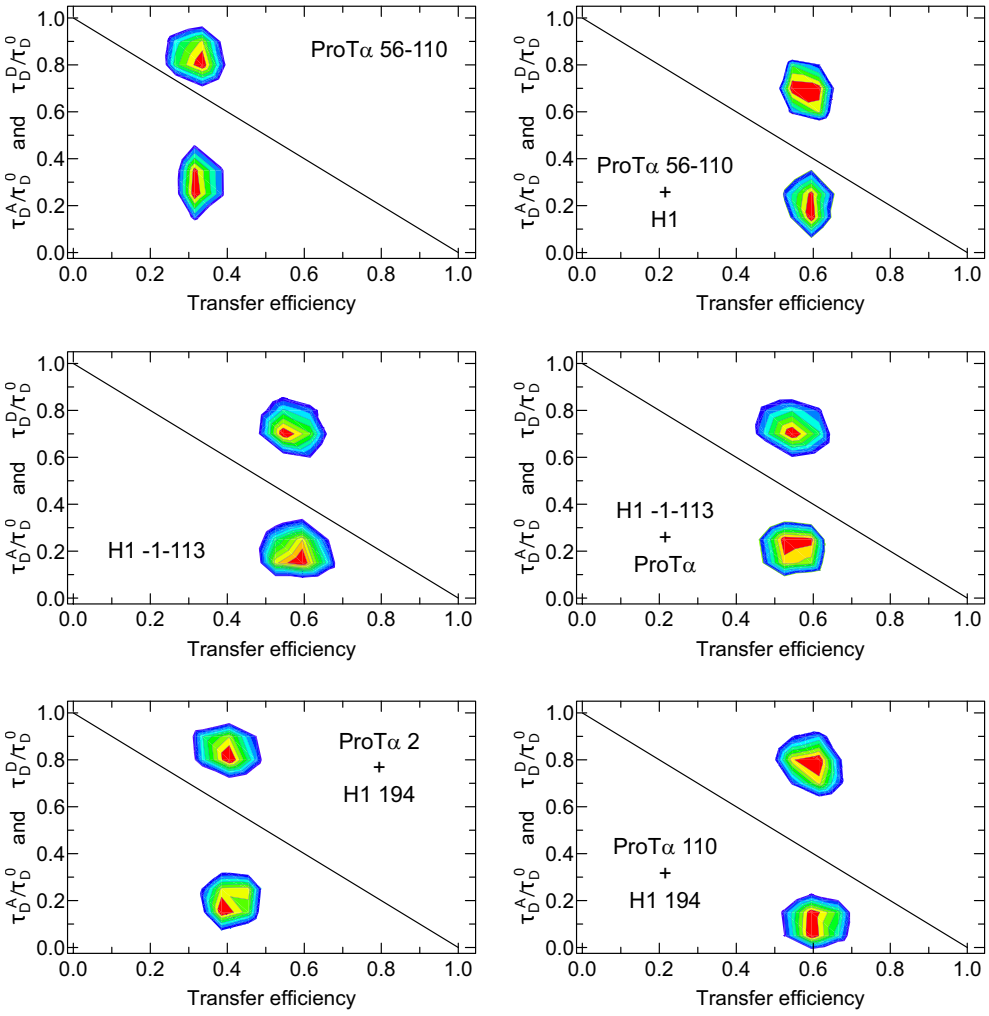


Extended Data Figure 3. Titration of ^{13}C,^{15}N-H1 with ProTα. (a) ^{1}H,^{15}N-HSQC spectra of free ^{13}C,^{15}N-GD (globular domain, dark green) and free ^{13}C,^{15}N-H1 (orange). The majority of the amide peaks of the GD overlap with the more dispersed peaks from full-length H1, indicating the similarity in structure of the GD in isolation and within H1. **(b)** Titration followed by ^{1}H,^{15}N-HSQC spectra of ^{13}C,^{15}N-H1 with 0- to 4-fold molar addition of ProTα. Data acquired on His_6-tagged H1. **(c)** CSPs relative to free H1 of eleven traceable H1 amide backbone peaks from the intrinsically disordered region (based on overlay with ^{1}H,^{15}N-HSQC spectra of GD (a)) upon 0 to 4-fold molar addition of ProTα plotted against concentration and times excess. Colors used for discriminability. **(d)** CSPs plotted against peak intensity ratios relative to the free state of H1 of the eleven H1 amides at 1× excess of ProTα. Colors as in (c). **(e)** Overlay of C^{α},H^{α} region from ^{1}H,^{13}C-HSQC spectra of free ^{13}C,^{15}N-H1 (blue) and ^{13}C,^{15}N-GD (green). The H1 ^{1}H,^{13}C-HSQC is dominated by intense clusters of peaks not present in the GD spectrum, consistent with the large fraction of repeats in the H1 disordered regions. **(f)** C^{α},H^{α} region of ^{13}C,^{15}N-H1 upon titration with ProTα. The lack of detectable changes in C^{α},H^{α} resonances is consistent with the absence of secondary structure induction in the disordered regions of H1 upon binding.

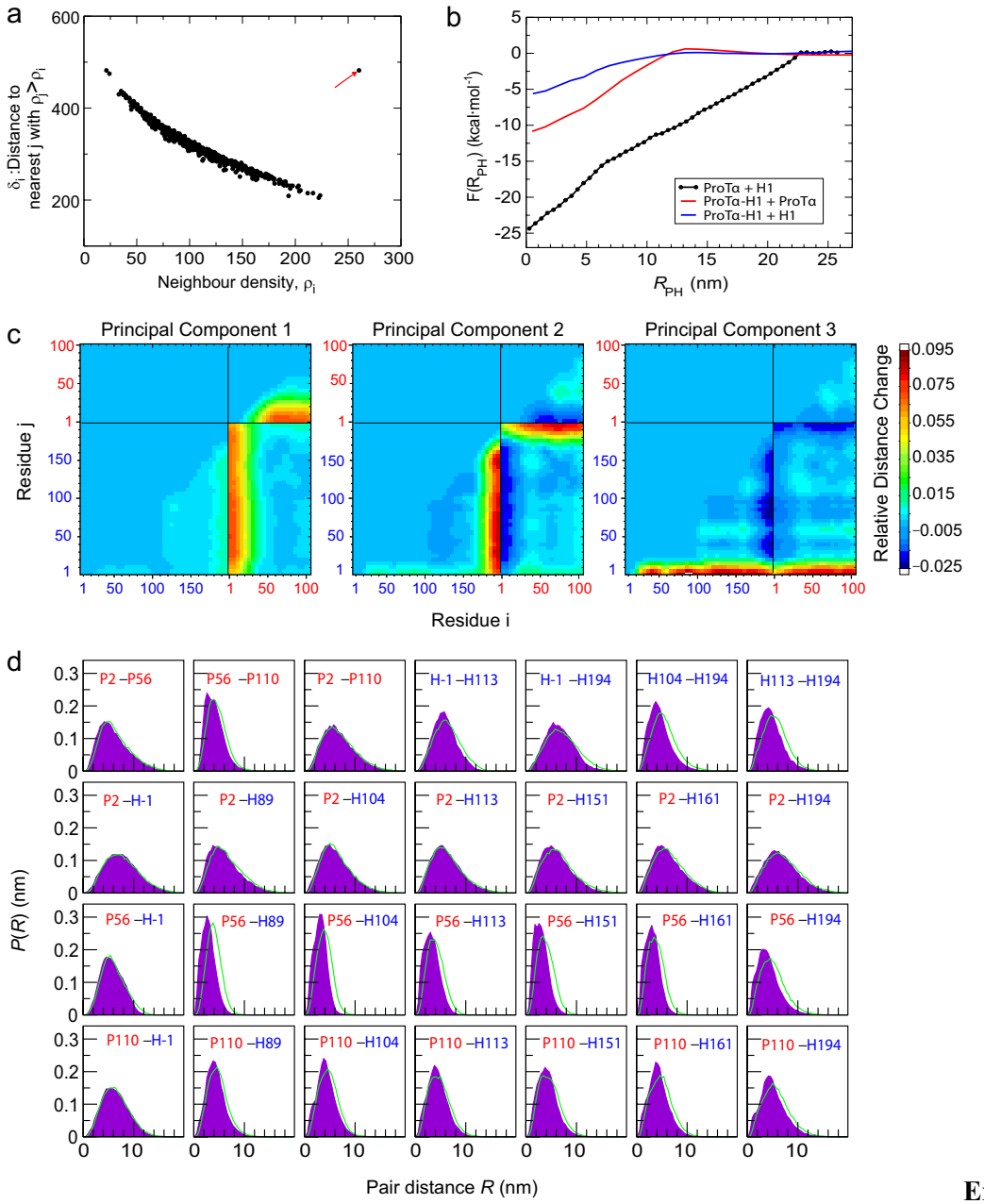


	D (m²/s)	R_H (nm)
^{15}N-ProTα	$5.5 \cdot 10^{-11} \pm 1 \cdot 10^{-12}$	2.89 ± 0.08
^{15}N-ProTα + H1	$4.7 \cdot 10^{-11} \pm 3 \cdot 10^{-12}$	3.4 ± 0.2
^{15}N-ProTα + GD	$5.0 \cdot 10^{-11} \pm 1 \cdot 10^{-12}$	3.20 ± 0.06

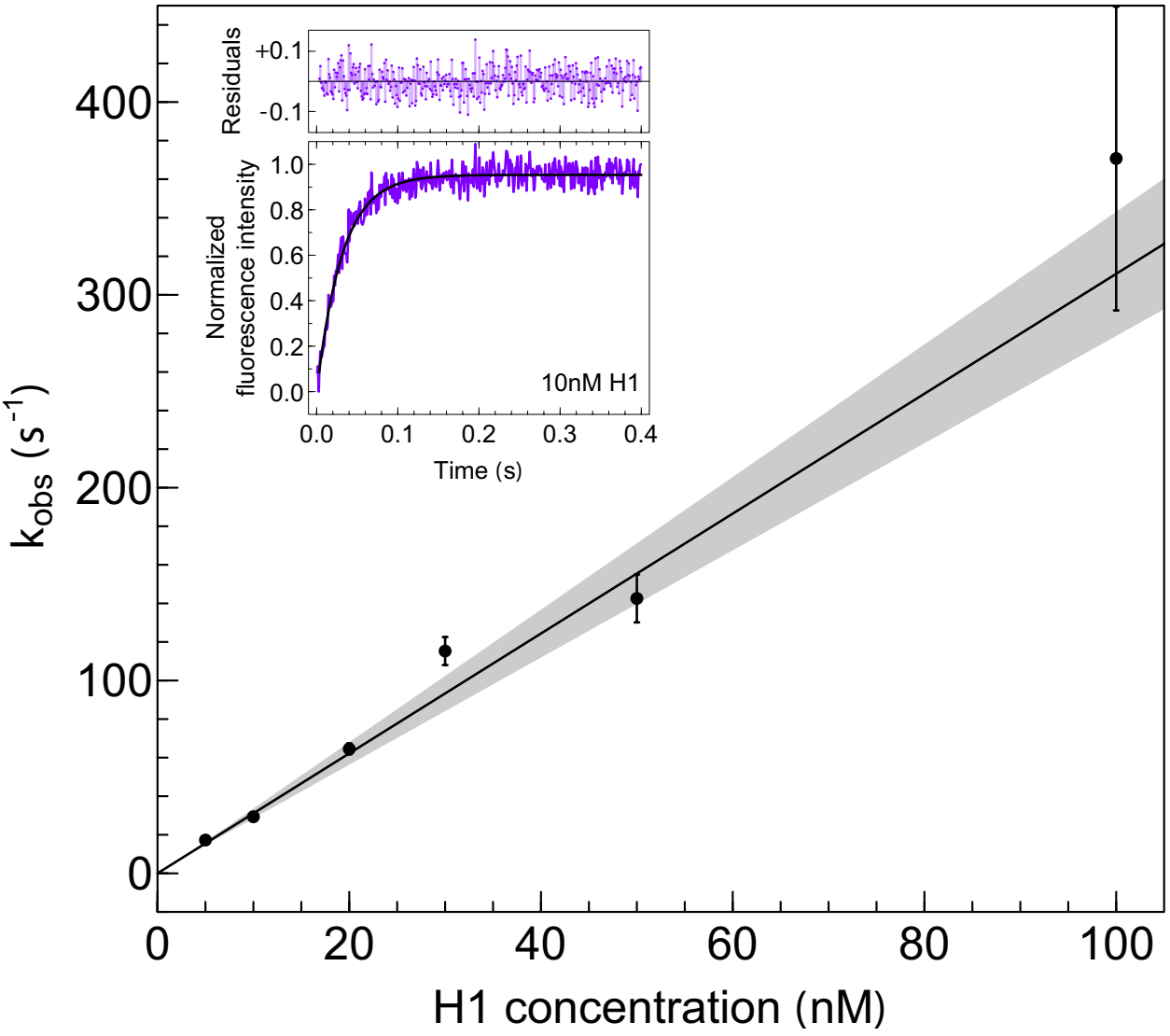
Extended Data Figure 4. Hydrodynamic radii and stoichiometry of the H1-ProTα complex. (a) Hydrodynamic radii, R_H, of free and bound ^{15}N-ProTα (100 μM) determined with pulsed-field gradient NMR at 283 K. The signal decays of free ^{15}N-ProTα (red), with H1 at a 1:1 molar ratio (purple), and with H1 GD at a 1:7 molar ratio (green) as a function of gradient strength, together with corresponding fits and a table of the diffusion coefficients and resulting R_H values. **(b)** R_H measured by 2f-FCS at 295 K. Lines show the R_H of H1 -1C (blue) and ProTα D2C (red) labeled with Alexa 594 in the absence of binding partner. Symbols show labeled ProTα (5 nM) in the presence of equimolar concentrations of unlabeled ProTα and unlabeled H1, with s.d.s indicated by error bars or shaded bands. **(c)** Stoichiometry ratio[71] versus transfer efficiency plots from intermolecular single-molecule FRET experiments with singly labeled protein variants as indicated in the panels. A stoichiometry ratio of 0.5 indicates a 1:1 complex. **(d,e)** Transfer efficiency changes at large excess of unlabeled binding partner for FRET-labeled ProTα C56C110 (d) and H1 C104C194 (e).



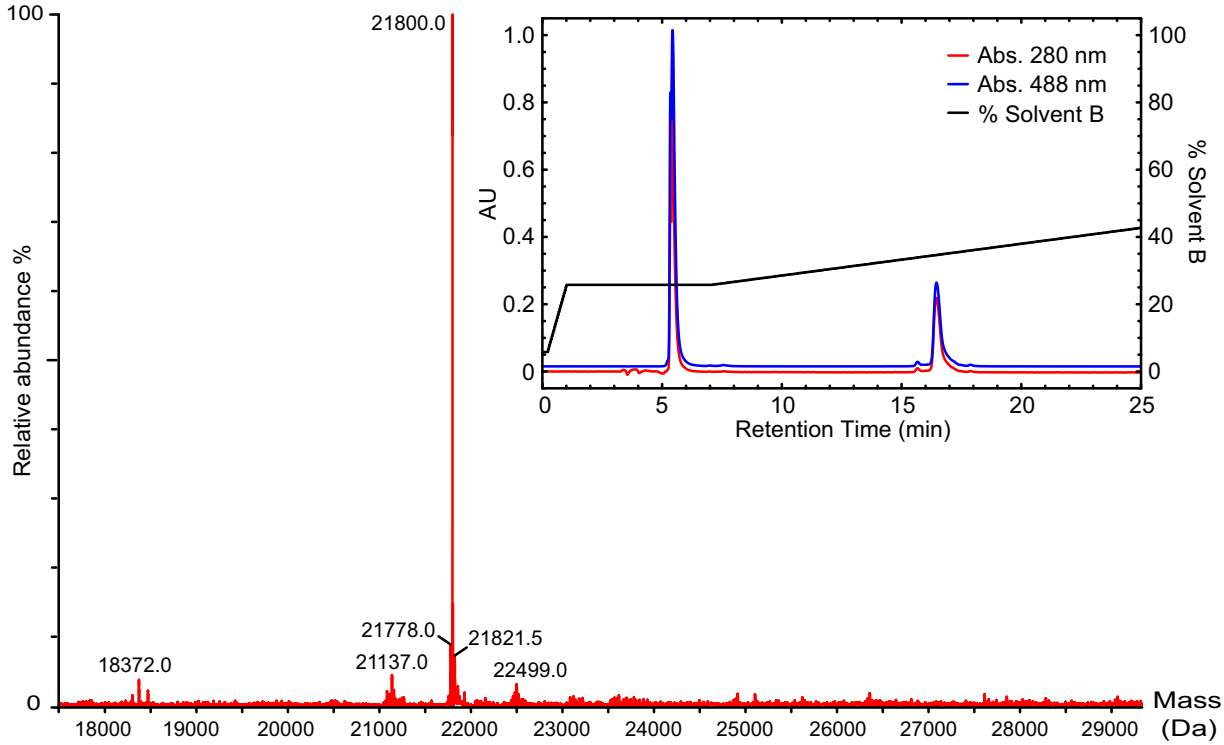
Extended Data Figure 5. Fluorescence lifetime analysis. Plots of the fluorescence lifetimes of donor (Alexa 488), τ_D^D, and acceptor (Alexa 594), τ_D^A, normalized by the intrinsic donor lifetime, τ_D^0, versus the ratiometric transfer efficiency, E (calculated from the number of donor and acceptor photon counts), as a diagnostic for the presence of a broad distance distribution rapidly sampled during the time of a fluorescence burst[28,33,34]. If fluctuations in transfer efficiency occur on a timescale between the donor fluorescence lifetime (~4 ns) and the burst duration (~1 ms), the normalized donor lifetimes cluster above, and the acceptor lifetimes below the solid diagonal line expected for a single fixed distance, as previously observed for intrinsically disordered proteins[34,72]. The large deviation from the diagonal observed for both unbound and bound ProTα and H1 supports the presence of broad, rapidly sampled distance distributions.



Extended Data Figure 6. Simulation results. (a) Decision graph using the Rodriguez-Laio clustering algorithm[70], showing only a single density maximum distant from other density maxima, i.e. a single distinct cluster. **(b)** Free energy for association of ProTα and H1 from simulation, yielding a K_D of 7 fM at $R_{PH} = 0$ (black curve). Blue and red curves are the free energies for addition of a second H1 or a second ProTα, respectively, to an existing H1-ProTα complex. **(c)** Principal component (PC) vectors shown as contact maps. Colors indicate the increase or decrease in each pair distance for that PC, relative to the other distances. ProTα and H1 residue numbers are indicated in red and blue, respectively. Each PC describes a feature of the chain arrangement: PC_1, e.g., captures the presence or absence of interactions between the ProTα N-terminus and H1. **(d)** Intramolecular (top row) and intermolecular (rows 2 to 4) distributions of distances corresponding to FRET labeling sites, for the ProTα-H1 complex (labels P*X*-H*Y* refer to residues *X* and *Y* in ProTα and H1, respectively). Filled distributions: simulations without explicit chromophores; green lines: simulations with explicit chromophores.



Extended Data Figure 7. Kinetics of H1-ProTα binding measured by stopped flow. FRET-labeled ProTα 56-110 is mixed rapidly with unlabeled H1 in TBS buffer, and the resulting increase in acceptor fluorescence is monitored (inset, measured at 10 nM H1 with single-exponential fit and residuals above, see Methods for details). Decay rates were obtained from single-exponential fits, assuming an instrument dead time of 3 ms. Standard errors for each H1 concentration were obtained via bootstrapping. The observed rates, k_{obs}, are shown as a function of H1 concentration (c_{H1}); for H1 concentrations between 10 and 100 nM, where pseudo-first order conditions apply (ProTα concentration after mixing was 2 nM), they were fit with $k_{obs} = k_{on}c_{H1} + k_{off} = k_{on}c_{H1} + k_{on}K_D$, using the independently determined K_D of 2.1 pM (Extended Data Table 2). The fit yields a bimolecular association rate coefficient of $k_{on} = (3.1 \pm 0.1)\cdot 10^9\ M^{-1}\ s^{-1}$ and an apparent dissociation rate coefficient of $k_{off} = (6.5 \pm 3.1)\cdot 10^{-3}\ s^{-1}$. The gray area represents the 95% confidence band.



Extended Data Figure 8. Example of the quality of the H1 preparation. Electrospray ionization mass spectrum of H1 T161C labeled with Alexa 488 (calculated mass 21,800 Da) and reversed-phase HPLC (Vydac C4) chromatogram (inset) showing absorption at 280 nm (red) and 488 nm (blue) and the elution gradient from solvent A (5% acetonitrile in H_2O + 0.1% TFA) to solvent B (100% acetonitrile) (black), illustrating the high purity of the sample. The peak at ~5.5 min corresponds to free Alexa 488, the peak at ~16.8 min to H1 T161C labeled with Alexa 488.

H1 (+53)	-1 23 CTENSTSAPAAKPKRAKASKKSTDHPKYSDMIVAAIQAEKNRAGSSRQSIQKYIKSHYKVGENADSQI 89 96 104 113 KLSIKRLVTTGVLKQTKGVGASGSFRLAKSDEPKKSVAFKKTKKEIKKVATPKKASKPKKAASKAPTK 151 161 193 KPKATPVKKAKKKLAATPKKAKKPKTVKAKPVKASKPKKAKPVKPKAKSSAKRAGKKKGGPR
H1 [a]**CTR** (+39)	103 SVAFKKTKKEIKKVATPKKASKPKKAASKAPTKKPKATPVKKAKKKLAATPKKAKKPKTVKAKPVKASK 193 PKKAKPVKPKAKSSAKRAGKKKGGPR
H1 [b]**NTR** (+18)	-1 GCTENSTSAPAAKPKRAKASKKSTDHPKYSDMIVAAIQAEKNRAGSSRQSIQKYIKSHYKVGENADSQI 113 KLSIKRLVTTGVLKQTKGVGASGSFRLAKSDEPKKSVAFKKTKKEI
GD (+9)	23 DHPKYSDMIVAAIQAEKNRAGSSRQSIQKYIKSHYKVGENADSQIKLSIKRLVTTGVLKQTKGVGASGS 96 FRLAK
ProTα (-44)	1 56 GPSDAAVDTSSEITTKDLKEKKEVVEEAENGRDAPANGNAENEENGEQEADNEVDEEEEEGGEEE 110 EEEEEGDGEEEDGDEDEEAESATGKRAAEDDEDDDVDTKKQKTDEDD

H1				**ProTα**				
Singly labeled		**Doubly labeled**		**Singly labeled**		**Doubly labeled**		
Alexa 488	Alexa 594	Alexa 488/ Alexa 594 (R_0=5.4 nm)[c]	Cy3B/ Abberior* 635 (R_0 = 5.9 nm)[c]	Alexa 488	Alexa 594	Alexa 488/ Alexa 594 (R_0 = 5.4 nm)[c]	Cy3B/ Abberior*635 (R_0 = 5.9 nm)[c]	Atto 550/ Atto 647N (R_0 = 6.6 nm)[c]
-1C, S89C, V104C, I113C, A151C, T161C,	-1C, S89C	-1C/113C, -1C/G194C, V104C/G194C, I113C/G194C	V104C/G194C	D110C	D2C, E56C, D110C	I113C/G194C, D2C/E56C, E56C/D110C, D2C/D110C	E56C/D110C	E56C/D110C

Extended Data Table 1. Sequences of protein constructs and fluorescently labeled variants of H1 and ProTα. (top) Sequences of H1 and ProTα wildtype and variants used. Bold yellow-shaded residues are positions mutated to Cys for fluorophore conjugation. Residues in red are part of protease recognition sites used to cleave the HisTag with thrombin (GGPR or GC) or HRV-3C (GP). (Note that the wt sequence of H1 starts with "T"; the preceding Cys residue (-1) was added for labeling.) The underlined H1 sequence indicates the globular domain (GD), identified based on a sequence alignment with the *G. gallus* homolog[20] (PDB access code 1HST, 82% sequence identity). Surface-exposed residues (as shown in Fig. 1a and 5b) are shaded in light blue. The net charge of each variant is indicated in parentheses. [a]C-terminal disordered region. [b]N-terminal disordered region including GD. **(bottom)** Labeled variants of H1 and ProTα. [c]Förster radius of the corresponding dye pair.

Affinity in TBS 165 mM (ProTα 56/110 Alexa 488/Alexa 594)				Affinity in TBS 205 mM		Intramolecular transfer efficiencies and distances in TBS 205 mM				
Ionic strength (mM)	K_D (nM)	H1 fragm.	K_D	ProTα	[d]K_D (nM)	ProTα	$E_{unbound}$	E_{bound}	$R_{unbound}$ (nm)	R_{bound} (nm)
[a]165	$(2.1^{+1.1}_{-0.8})\cdot10^{-3}$	[b]CTR	40^{+6}_{-4} pM	D2C/D110C Alexa 488/594	2.0 ± 0.13	E56C/D110C Alexa 488/59	0.36	0.54	$7.5^{+1.0}_{-0.3}$	$5.8^{+0.7}_{-0.1}$
180	$(37 \pm 5)\cdot10^{-3}$					E56C/D110C Cy3B/Abb.*635	0.41	0.56	$7.6^{+0.6}_{-0.5}$	6.2± 0.4
205	1.0 ± 0.1	[b]NTR	173^{+29}_{-28} nM	E56C/D110C Cy3B/Abb.*635	1.0 ± 0.10	E56C/D110C Atto 550/647N	0.45	0.59	$8.1^{+0.6}_{-0.5}$	$6.7^{+0.5}_{-0.4}$
240	26 ± 3			E56C/D110C Atto 550/647N	3.1 ± 0.20	V104C/G194C Cy3B/Abb.*635	0.18	0.52	$11.6^{+1.8}_{-1.2}$	$6.5^{+0.5}_{-0.4}$
290	$(2.5 \pm 2.1)\cdot10^{2}$	[b]GD	$1.9^{+0.3}_{-0.3}$ µM[c]	H1		H1	In TBS 165 mM			
						D2C/D110C Alexa 488/594	0.18	0.33	$10.6^{+1.6}_{-1.1}$	$7.9^{+0.7}_{-0.6}$
330	$(1.4 \pm 0.4)\cdot10^{3}$			V104C/G194C Alexa 488/594	3.5 ± 0.23	I113C/G194C Alexa 488/594	0.23	0.58	$9.5^{+1.1}_{-0.9}$	5.5± 0.4
340	$(4.0 \pm 1.8)\cdot10^{3}$					V104C/G194C Alexa 488/594	0.14	0.52	$11.4^{+2.1}_{-1.4}$	5.9± 0.4

ProTα A-594 / H1 A-488	τ_r (ns) D2C	τ_r (ns) E56C	τ_r (ns) D110C	Labeled Protein (Alexa 488/ Alexa 594)	τ_r (ns) unbound	τ_r (ns) bound
H1-1C	180^{+19}_{-16}	191^{+22}_{-19}	169^{+19}_{-16}	ProTα E56C/D110C	29±2	102^{+3}_{-2}
H1 I113C	121^{+13}_{-11}					
H1 A151C	124^{+13}_{-12}	98^{+16}_{-2}		ProTα D2C/D110C	33±2	66±2
H1 G194C	156^{+16}_{-14}		142^{+19}_{-15}			
A-488 / A-594				ProTα D2C/D110C	78^{+15}_{-9}	133^{+10}_{-7}
H1 G194C			120^{+13}_{-12}	H1	118^{+24}_{-14}	143^{+5}_{-4}

Extended Data Table 2. Binding affinities, molecular dimensions, and reconfiguration times of fluorescently labeled H1 and ProTα. (top left) Affinities of labeled ProTα for H1 at different ionic strength (IS) and for H1 fragments for 165 mM IS ([b]see Extended Data Table 1). Uncertainties for the IS dependence are standard errors estimated from two independent titrations ([a]uncertainty at 165 mM: see Methods), for fragment binding from dilution errors (see Methods). [c]Apparent K_D from fraction of all bound species. **(top center)** Binding affinities of ProTα and H1 labeled with different dye pairs for the respective unlabeled partner. [d]Uncertainties based on dilution errors. **(top right)** Transfer efficiencies and average distances of ProTα and H1 labeled with different dye pairs in the bound (R_{bound}) and unbound state ($R_{unbound}$). Uncertainties in distance are based on an estimated systematic error of ± 0.05 in the transfer efficiency from instrument calibration for the different dye pairs. **(bottom left)** Intermolecular reconfiguration times for the complex of donor-labeled H1 and acceptor-labeled ProTα and *vice versa*. **(bottom right)** Reconfiguration times of doubly labeled ProTα and H1 (unbound and bound). Uncertainties estimated by propagating the error on the transfer efficiency (± 0.05).

Extended Data References

71 Kapanidis, A. N. *et al.* Alternating-laser excitation of single molecules. *Acc. Chem. Res.* **38**, 523-533 (2005).

72 Hoffmann, A. *et al.* Mapping protein collapse with single-molecule fluorescence and kinetic synchrotron radiation circular dichroism spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 105-110 (2007